

**APPLICATION FOR  
UNITED STATES PATENT  
IN THE NAME OF**

**JAMES L. HEPWORTH AND FAISAL SHAH**

**ASSIGNED TO**

**eMARKMONITOR INC.**

**FOR**

**METHOD AND SYSTEM FOR SEARCHING AND MONITORING  
INTERNET TRADEMARK USAGE**

**DOCKET NO.: 23802-250800**

**Prepared By:**

**PILLSBURY MADISON & SUTRO LLP  
725 South Figueroa Street, Suite 1200  
Los Angeles, CA 90017  
Telephone (213) 488-7100  
Facsimile (213) 629-1033**

**Express Mail No.: EJ 820 109 882 US**

TITLE OF THE INVENTION

METHOD AND SYSTEM FOR SEARCHING AND MONITORING INTERNET

TRADEMARK USAGE

5 RELATED U.S. APPLICATION DATA

This is a continuation-in-part of application serial No. 09/359,924, filed July 22, 1999.

BACKGROUND OF THE INVENTION

1. Field of the Invention

09/359,924 - 07/10/00  
15  
20  
The present invention provides a process and an apparatus for: (1) comprehensively searching the Internet for uses of trademarks, tradenames, celebrity and/or famous names; (2) comprehensively searching the Internet for potentially infringing uses of trademarks in domain names and other Web page content as it relates to homonym and phonetic equivalents; (3) comprehensively searching the Internet for potentially infringing uses of trademarks, tradenames, celebrity and/or famous names as it relates to homonyms and phonetic equivalents; and (4) comprehensively searching the Internet for uses of trademarks, tradenames, celebrity and/or famous names as they relate to images.

2. Description of the Related Art

20  
The rapid rise of the Internet has created an explosion of information and commercial speech and with it an increase of intellectual property issues, particularly trademark issues. The earliest form of this issue was (and is) the conflict between domain name rights and trademark rights. Under our present system of registration, any party ("Registrant") can register with

InterNIC (a federal government organization established to register the domain name database) to own any domain name, regardless of trademark ownership, affiliation or association with such name. As a result, Registrants would register hundreds of domain names with InterNIC with the primary goal of selling these same names back to the individuals and/or companies that actually own the same registered trademark. These types of Registrants have been referred to as "Cybersquatters". While a loose mechanism exists in the form of a dispute resolution policy to dispute the "land rush" domain name claims, there still is a need to understand and monitor how trademark rights (either registered in a particular country or based upon common law rights) are being enforced and used on the Internet.

The Internet is mostly comprised of, among other things, the Web, Newsgroups, and FTP. The Web is comprised mostly of Web pages that are constructed using HTML. This language is written using codes or tags which allow a Web page to be viewed in a coherent or logical fashion. This is also the programming language that allows a user to "jump" or "hyperlink" from one Web page to another with the click of the mouse. These codes or tags are hidden from view from the individual user viewing the Web page. Also hidden from view are image tags, hidden text, titles and URLs of the Web page. The reason the user does not see this hidden tag, text, title and URL is so that the user can interface with the Web in a "point and click" fashion. Presently, there are over 500 million Web pages on the Internet and this number is growing exponentially each year.

Search engines (e.g., Yahoo!, Lycos, Infoseek, etc.) are Internet portals that catalog Web sites on the Internet in a "search, store and retrieve fashion". First, the search engine searches or "crawls" the Internet and gathers available Web pages on the Internet. The Web pages are then stored and cataloged in that particular search engines' database based on certain criteria (i.e.,

how frequent keywords appear in either the hidden text, text, meta-tag, title or URL of the Web page). This function enables the search engine to then retrieve a relevant indexed Web page when a keyword query is entered by an end user. Search engines determine relevancy by giving different values to the above criteria (i.e., one search engine may index Web pages in accordance with how many times a keyword appears in the hidden text of a Web page while another search engine may index its Web pages in accordance with how many times a keyword appears in the text of a Web page). This is why when a keyword is entered on one search engine, this search engine may return results and rank Web sites differently than on other search engines. Some search engines also rank other Web Sites based on criteria such as link popularity (i.e. how many Web Sites are linking to that particular Web Site based on a certain keyword). Drawing large amounts of Internet traffic is the primary goal of most if not all Web sites. It is therefore critical that a Web site be easily located and be ranked as high as possible on an end user's search result list.

Companies and individuals have taken advantage of these programmed searches performed by search engines by inserting registered trademarks, tradenames, celebrity and/or famous names as meta-tag keywords or as hidden words within the Web page itself. In addition to hidden text or words within the Web page itself, a Web site may also contain registered trademarks, tradenames, celebrity and/or famous names in the URL or the title of the Web page, and may contain a reference to an image (in the form of a trademark, tradename, celebrity and/or famous name), which image can be viewed by the end user. By using the registered trademarks, tradenames, celebrity and/or famous names in the hidden text, meta-tags, titles, URLs, and image text, third-party Web sites are able to cleverly draw traffic away from authorized Web pages. For example, assuming a search engine indexes its Web pages in accordance with how frequent a

keyword appears in the meta-tag of a Web page and assuming a registered trademark that is owned by party A is listed in a meta-tag numerous times in an unrelated Web site run by party B, it is more than likely that if an unknowing user of the Internet enters a query on a standard search engine seeking all Web sites related to that particular trademark, then, the search report will

5 return a list of Web sites with the unrelated site of party B as the top site among such Web sites. This result confuses the end user in that he or she believes that the first listed Web site of party B may, in fact, be endorsed or associated to party "A", the actual owner of the registered trademark.

As discussed above, search engines typically index Web sites in their database in accordance with several factors, including, but not limited to, the frequency of the number of keywords appearing in the hidden programming code and text of a Web site. Web site programmers take advantage of this indexing method by the search engines by programming their Web site's HTML source code to include irrelevant keywords in their Web site's title, meta-tags, text, and hidden text. As a result, a keyword search query entered into a search engine may yield results that place the irrelevant site higher on the list of search results than the company's authorized site which the user actually desires to find. This result can divert valuable traffic away from the authorized site.

The need to draw high levels of Internet traffic to one's Web site has resulted in the practice of "Spamdexing". This is a technique used by Web site owners to bring their Web site

20 to the top of a list of search results. The practice involves repeating keywords over and over in text usually at the top of a Web page and/or at the bottom of a Web page in very small letters. It also involves repeating keywords in meta-tags, addresses, hidden text, titles or URLs. These keywords are frequently registered trademarks. If a Web site is able to successfully "spamdex"

5

## SUMMARY OF THE INVENTION

(a) providing the trademark, tradename, celebrity or famous name to be searched;

(c) searching the Internet with a plurality of search engines with the search string that identifies the target URL, meta-tags, hidden text, text, titles, domain names and images in order to identify the target URL with hits;

(e) providing the search results consisting of identified hits and location within each  
n name.

-5-

selected from the group consisting of Yahoo!, LookSmart, Mining Co., Goto, Thunderstone, Dogpile Open Directory, Infoseek, Infoseek Ultra, Excite, Excite Guide, Alta Vista, Lycos, Lycos A2Z Directory, Webcrawler, What U Seek, Magellan, NetGuide, Galaxy, Open Text, HotBot, Point, REX, Northern Light, AskJeeves, Go, Inktomi Powered Engines and  
5 combinations thereof.

The present invention further provides an apparatus for searching for uses of trademarks, tradenames, celebrity and famous names on the Internet for a client, comprising:

(a) a plurality of database servers wherein the database servers store information of searching activity, including the identity of the client and identity of domain names and URLs having discovered hits relating to a trademark, tradename, celebrity or famous name or homonyms or phonetic equivalents related thereto;

(b) a plurality of crawler servers wherein the crawler servers comprise a search engine, and wherein the crawler servers are connected to the database servers and the Internet and function to crawl Internet sites and Web pages identified by previous hits from search engines, crawlers or database servers; and

(c) one or a plurality of user servers that comprise a program for determining a plurality of homonyms and phonetic equivalents of a particular trademark, tradename, celebrity and/or famous name to be searched.

Preferably, the apparatus further comprises a means for encryption of communication  
20 between the user server and the client. Preferably, the client comprises a server that will communicate with the user server(s) by means of an Internet Web site.

According to an embodiment of the present invention, a method of searching and reporting the incidence of at least one trademark, tradename, celebrity name, and/or famous

name on the Internet includes the step of providing at least one trademark, tradename, celebrity name, and famous name to be searched. A search string is created that includes at least one trademark, tradename, celebrity name and famous name. A URL address is provided of a Web page on the Internet to be searched. Next, the Web page is accessed and searched for hits

5 corresponding to the search string. Then, the search results of identified hits are provided within the Web page.

According to another embodiment of the present invention, a system for searching and reporting the incidence of at least one trademark, tradename, celebrity name, and/or famous name on the Internet, includes a computer system having a connection to the Internet and a software program executing on the computer system. The software program is adapted to receive an input of at least one trademark, tradename, celebrity name, and famous name to be searched, to create a search string including at least one trademark, tradename, celebrity name, and famous name, to receive a URL address of a Web page on the Internet to be searched, to access and search the Web page for hits corresponding to the search string, and to provide search results of identified hits within the Web page.

According to yet another embodiment of the present invention, a software program executing on a computer system for searching and reporting the incidence of at least one trademark, tradename, celebrity name, and/or famous name on the Internet, includes instructions for receiving at least one trademark, tradename, celebrity name, and famous name to be

20 searched. The software program has instructions for creating a search string including at least one trademark, tradename, celebrity name, and famous name. Instructions are provided for receiving a URL address of a Web page on the Internet to be searched. The software program also includes instructions for accessing and searching the Web page for hits corresponding to the



search string. Instructions are also included for providing search results of identified hits within the Web page.

#### BRIEF DESCRIPTION OF THE DRAWINGS

5           Fig. 1 illustrates a diagram of the inventive apparatus showing the interconnections and communications between the user server(s), crawler servers, database servers and the Internet;

          Fig. 2 shows a search flow diagram for listing each step of a hypothetical search using the inventive process;

          Figs. 3A to 3C illustrate an example of a basic search report generated utilizing a variety of commercially-available search engines;

          Fig. 4 illustrates a diagram of the components of a keyword matching system for searching a Web page according to an embodiment of the present invention; and

          Fig. 5 illustrates a flow chart diagram for searching for keyword matches in the programming code of a Web page according to an embodiment of the present invention.

#### DETAILED DESCRIPTION

As will be made clear below, the general availability of personal computers and of the Internet and the advent of the present invention make possible mechanisms to enable and encourage legal professionals and other end users to monitor potential trademark and tradename infringement and the unauthorized use of celebrity and/or famous names for commercial gain.

## DEFINITIONS

Browser is a client program (software) that is used to look at various kinds of Internet resources.

5 Crawler is a software application that indexes Web pages and other Internet sites by public IP address or top level domain name (e.g., Nike.com).

10 Domain name is a unique name that identifies an Internet site. Domain names always have two or more parts, separated by dots. The part on the left is the most specific and the part on the right is the most general. A given machine may have more than one domain name but a given domain name points to only one machine. For example, the domain names "matisse.net"; "mail.matisse.net"; and "workshop.matisse.net" can all refer to the same machine, but each domain name can refer to no more than one machine.

15 FTP (file transfer protocol) is a common method of moving files between two Internet sites. FTP is a special way to login to another Internet site for the purposes of retrieving and/or sending files.

20 Hidden text is HTML text that is coded with tags to prevent the text from being visible to the end user.

HTML or Hypertext Markup Language is the coding language used to create hypertext documents for use in the Web. HTML looks like typesetting code whereby a block of text is surrounded by codes that indicate how it should appear. In HTML one can specify that a block of text or a word be linked to another file in the Internet.

Hypertext is text that contains commands that communicate with the Browser and tells the Browser how to display text, graphics and multimedia files.

Hit occurs any time a piece of data matches a term in a search string. One example would be the results that are returned from a Yahoo! or other search engine.

Internet is a global speaking computer network of networks whereby the computer networks communicate with each other pursuant to certain protocol. Tens of thousands of such computer networks exist, ranging from university networks to corporate local area networks to large online services (i.e., America Online).

Meta tag is part of the head or top of an HTML document and provides information that describes the document in various ways. It contains information for search engines to use in adding Web pages to their search indexes. It can also be used to search locally for similar files or files that need reviewing or updating. Information in each <META> tag is expressed as a NAME = and value =pair. The NAME can be used to distinguish one type of <META> statement from another.

Newsgroups is the name for discussion groups on the Internet.

Server is a computer and software package that provides a specific kind of service to a client software running on other computers. A single server machine can have several different server software packages running on it and providing many different servers to clients on a network.

SSL (secure sockets layer) is a protocol designed to enable encrypted, authenticated communications across the Internet. SSL is used mostly in communications between Web Browsers and Web Servers. URLs that begin with "http" indicate that an SSL connection will be used. SSL provides privacy, authentication and message integrity. In an SSL connection, each side of the connection must have a security certificate, which each side's software sends to the other. Each side then encrypts what it sends using information from both its own and the other

side's certificate, ensuring that only the intended recipient can de-crypt it and that the other side can be sure the data came from the place it claims to have come from, and that the message has not been tampered with.

Title is the name in HTML which is given to the name of a particular HTML document

5 (e.g., <TITLE>Pepsi@HomePage</TITLE>).

Text is the words, sentences and paragraphs on a Web page that is visible to the end-user.

URL (uniform resource locator) is the standard way to give the address of any resource on the Internet that is part of the World Wide Web (e.g., <http://www.microsoft.com> or <ftp://ftp.microsoft.com>).

Usenet is a world-wide network of discussion groups, having comments passed among hundreds of thousands of machines. Not all Usenet machines are on the Internet as it is decentralized with many discussion areas called Newsgroups.

Web page is an HTML document that resides on the Web and incorporates text, graphics, sound, and other multimedia elements.

World Wide Web (Web) is a globally-connected network that allows text, graphics, sound files, etc. to be mixed together.

## NETWORK ARCHITECTURE

The components of a preferred embodiment of the present invention are illustrated in Fig.

- 20 1. A standard personal computer, computer workstation or computer server with adequate processing power and memory may be used as a User server 10. In one embodiment it operates as a Web server, receiving and responding to client 14 requests for access to the inventive process services. User servers 10 must be capable of rapidly performing database queries as well

as handling input and output needs. A Sun Microsystems® 300 MHz UltraSparc® II processor is one example of a CPU that can be used. A similar processor such as a 550 MHz Compaq® Alpha 21164 or 550 MHz Intel® Inc. Pentium® III may also be used. These processors are listed only as examples, and any suitable processor device may be utilized.

5 Database servers 11 provide a means for storing searches and for accessing such previous searches for determining a history of use of a particular trademark or tradename. The present Web site comprises a database server 11 that stores all prior searches indexed by user identification and search strings so that prior searches can be followed up. The database servers 11 are connected to the search engine servers 12 through the Internet 15, to the user servers 10 and potentially communicating to the client 14 through a Web site operated by the user servers 10. Data storage device(s) used in the database servers 11 consist of storage media such as hard disk magnetic storage, magnetic tape, or CD-ROM drives. Data storage device(s) consist of databases used in the processing of transactions in the present invention.

## PROCESS

Fig. 2 shows a search flow diagram for listing each step of a hypothetical search using the inventive process. In step 1, the user enters the search terms and search name for stored search results in the user interface and clicks on "Launch This Search." The search string is sent to the user database, which checks the user input, and returns the proper error messages if needed. If  
20 there are no errors on user input then the user is presented with a page stating: "Your search has been launched, we will e-mail you once it is complete." The search term(s) is then compared to a homonym and phonetic (word dictionary) database and concatenated to the search text and stored in the user database to form the search string. See step 2.

The search string is stored in the user database and queued for the next scheduled “universal” search done by the search engine. See step 3.

The inventive process retrieves the search string from the user database and searches the Internet and other search engines. See step 4.

- 5           The results of the search are processed and stored into a Results Database or Matrix. Each entry in this database contains mainly the URL of the found pages. See step 5.

The crawler retrieves the URL from the Results Database or Matrix and crawls to the retrieved URL to retrieve the complete Web page. The crawler then processes the returned page, parses the required information and stores the result into an Index Archive. See step 6.

When the search for this search string is completed, the results are fed to the Report Engine. See step 7.

The Report Engine formats the report in the user’s preferred report format and is stored in the Users database under the User’s profile. See step 8.

The present invention automatically determines homonyms and phonetic equivalents by a reverse spell check lookup. Examples of reverse spell-check lookup include “correct word is like” and reverse spell check lookup would also include “loke”, “liki”, “lich” and “lick”. In MS Word, for example, one can right click on a word and obtain a list of reverse spell check words. Phonetic examples would include “lyk” and “lyke”. Phonetic equivalents of “light” would include lite and lyte and lit. Phonetics may be looked up in an on-line dictionary.

- 20           In a present embodiment of the invention, the initial search first uses existing search engines to identify hits of the searched text homonyms and phonetic equivalents to form a search string as a first step. However, the results of such a search do not indicate where in the Web site (e.g. Web page) the identified term is located. Therefore, the inventive process uses a crawler in

a second step to go to each identified URL where a hit has been located on a search and separately index those identified URLs for the location(s) of the terms in the search string.

Figs 3A-3C illustrate an example of a basic search report generated utilizing a variety of commercially-available search engines. The print out shows the results of hits broken down by search engine used in terms of the number of hits and the URLs identified. The crawler then goes to each URL identified with a hit and breaks down the hit locations according to meta tag, hidden text, text, title, domain name, and images. Therefore the report will indicate where on each URL or domain, the hits are located. In a preferred embodiment, the report further prints out the actual text, hidden text, title, domain name, meta-tag, or image with the search term(s) highlighted such that the report can be used as evidence for a request to cease infringing activity.

The present invention further provides the ability to determine the ranking of a Web site according to each search engine. This allows a registered trademark owner to determine in what position an authorized Web Site is ranked in each search engine. Each search engine has different search parameters for ranking. In one embodiment, the search engine provides the user with the ranking regardless of the parameter that the search engine uses. This search uses a customized CGI (Common Gateway Interface) script (standard Internet script) by having it return the rank and URL of each Web site. The crawler is then launched such that the crawler then crawls each Web page to determine the location on the Web page of the search terms of the search string.

The present invention is further able to determine the name and address of the owner of a particular domain name by automatically accessing the InterNIC or registrar "whois" database, wherein the summarized report contains the name and address of each owner of each matching domain name.

A search can be customized in its report format for everything that can be reported. For example, a basic report is illustrated in Figs. 3A-3C. The report may show the unauthorized use of the "markfamus" trademark in the URLs and/or the meta-tags, which would divert traffic that was intended to go to an authorized "markfamus" site to that of an unauthorized Web site. The report can further identify those unauthorized URLs and obtain the name and address of the owners of such URLs and even print out pages from the unauthorized URLs as proof of unauthorized uses that can be appended to a cease and desist letter sent to the owners of such unauthorized URLs. In addition, the database server keeps track of previous searches and can regularly update any changes made to offending URLs. Therefore, the inventive process provides a valuable Internet-based tool for trademark and tradename owners and for those connected with celebrity and/or famous names to monitor and police their intellectual property rights on the Internet.

In addition, the present invention provides a mechanism for encrypting data and other information crossing between the client/customer of the inventive process service and the server. Such mechanisms for encryption used on the Internet include SSL's and a Verisign server 13.

Fig. 4 illustrates a diagram of the components of a keyword matching system for searching a Web page according to an embodiment of the present invention. Similar to the system discussed above, the client computer 14 may be connected to the Internet 15 for remote access to the keyword matching system. The keyword matching system is adapted to receive from the user at the client computer 14 a URL address of a Web page to be searched on an electronic media server 20 across the Internet 15. The keyword matching system preferably includes a parsing server 22, which performs keyword matching on the Web page that is retrieved from the URL address provided by the user. A database server 11 may also be



incorporated to access a database 26 of homonyms and phonetic equivalents to the keyword(s) being searched so as to additionally search for any such homonyms and phonetic equivalents that may exist within a Web page. The homonyms and phonetic equivalents may be transmitted to the parsing server 22 to allow for the additional search for homonyms and phonetic equivalents within a retrieved Web page. A Web/electronic formatting server 24 may be provided to take the results obtained from the parsing server 22 and present the results in a user-friendly format to the user at the client 14, via a Web browser. All of the servers discussed herein regarding the keyword matching system may be embodied within a single server (computer system), or each individual server discussed above may be each comprised of multiple computer systems.

Fig. 5 illustrates a flow chart diagram for searching for keyword matches in the programming code of a Web page according to an embodiment of the present invention. For example, a Web site may be used for hosting a keyword matching system. The keyword may be a trademark, tradename, celebrity name, and/or famous name to be searched. Although the primary use of the present invention is to search for and monitor trademark-type usages, any other keyword or combination thereof may be searched as well. Using the keyword matching system of the present invention, the user may simultaneously search for matching keywords in the programming code (such as hidden text and meta-tags) and the visible text of a Web page, which information may be later categorized and summarized for the user.

First, the user may remotely connect to the keyword matching system via the Internet, as in step 400. Once the connection is made, the keyword matching system prompts the user to enter a known URL address and a specific keyword or keywords for which the user is searching within the Web page at the URL address. Instead of using a search engine to locate the URL addresses of potentially infringing Web pages, the keyword matching system in Fig. 4 allows the

user to input a specific known URL address for a Web page to be analyzed. Upon entry of the URL address and the keyword(s) by the user or any automated process, the keyword matching system accesses the Web page on the Internet with the provided URL address, and subsequently searches the source code (HTML) of the Web page for the requested keyword(s) 410, including conducting a search in the meta-tag section, the hidden text, the visible text, the title, and any hyperlinks within the Web page. Additionally, the search may be conducted of the image text as well, that is, the names of the images on the Web site, such as the names of the image files themselves, or the names or descriptions of the images assigned in the Web page source code. The search string, which may consist of just a single keyword, used to search for any matches within the Web page code may further include a plurality of other words, such as homonyms and phonetic equivalents (as discussed above) to the keyword(s) to be searched 430. If no keywords are located in the Web page, then a message is displayed to the user reporting the same 420.

Once the Web page has been retrieved and analyzed for keyword matches (including any homonyms and phonetic equivalents), the keyword matching system may provide a report 440 to the user by highlighting the keyword(s) and providing the location of the keyword within the Web page, such as, for example, in the title, the meta-tags, the text, the hidden text, the hyperlinks, etc. The keyword matching system may further summarize the number of matches or "hits" within each of these different categories.

While the description above refers to particular embodiments of the present invention, it will be understood that many modifications may be made without departing from the spirit thereof. The accompanying claims are intended to cover such modifications as would fall within the true scope and spirit of the present invention. The presently disclosed embodiments are therefore to be considered in all respects as illustrative and not restrictive, the scope of the

invention being indicated by the appended claims, rather than the foregoing description, and all changes that come within the meaning and range of equivalency of the claims are therefore intended to be embraced therein.

09612766-071000